

# The minimum entropy output of a quantum channel is locally additive

Gilad Gour<sup>1,\*</sup> and Shmuel Friedland<sup>2,†</sup>

<sup>1</sup>*Institute for Quantum Information Science and Department of Mathematics and Statistics,  
University of Calgary, 2500 University Drive NW, Calgary, Alberta, Canada T2N 1N4*

<sup>2</sup>*Department of Mathematics, Statistics and Computer Science University  
of Illinois at Chicago, 851 S. Morgan Street, Chicago, IL 60607-7045*

We show that the minimum von-Neumann entropy output of a quantum channel is locally additive. Hastings's counterexample for the additivity conjecture, makes this result quite surprising. In particular, it indicates that the non-additivity of the minimum entropy output is a *global* effect of quantum channels.

PACS numbers:

## I. INTRODUCTION

One of the most fundamental questions in quantum information concerns with the amount of information that can be transmitted reliably through a quantum channel. Despite of the significant progress in recent years [1, 2, 4, 5, 7, 11, 14, 15, 17, 19–21, 25, 26, 26, 27], as pointed out in [4], this question remained surprisingly wide open. The main reason for that is related to the additivity nature of the classical or quantum capacities of quantum channels to transmit information [15]. Recently, it was shown that both the Holevo expression for the classical capacity [14] and the quantum capacity [27] are not additive in general. The additivity of the Holevo expression for the classical capacity was an open problem for more than a decade and was shown by Shor [26] to be equivalent to three other additivity conjectures; namely, the additivity of entanglement of formation, the strong super-additivity of entanglement of formation, and the additivity of the minimum entropy output of a quantum channel.

In [14] Hastings gave a counterexample to the last of the above additivity conjectures and thereby proved that they are all false. Hastings counterexamples (see also [5]) exist in very high dimensions and an estimate of these extremely high dimensions can be found in [11]. Earlier, in [26], Shor pointed out that if the additivity conjectures were true, perhaps the first step towards proving them would be to prove local additivity. We show here that this local additivity conjecture is indeed true, despite the existence of counterexamples to the original additivity conjectures. Our results therefore demonstrate that the counterexamples to the original additivity conjecture exhibit a global effect of quantum channels.

As we pointed out in Appendix B of [10], both the local and global additivity conjectures are false over the real numbers. This in turn implies that a straightforward argument involving just directional derivatives could not provide a proof of local additivity in the general complex case. Hence, to show local additivity we use strongly the complex structure.

In quantum information theory, quantum channels are the natural generalizations of stochastic communication channels in classical information theory. They are described in terms of completely-positive trace preserving linear maps (CPT maps). A CPT map  $\mathcal{N} : H_{d_{\text{in}}} \rightarrow H_{d_{\text{out}}}$  takes the set of  $d_{\text{in}} \times d_{\text{in}}$  Hermitian matrices  $H_{d_{\text{in}}}$  to a subset of the set of all  $d_{\text{out}} \times d_{\text{out}}$  Hermitian matrices  $H_{d_{\text{out}}}$ . Any finite dimensional quantum channel can be characterized in terms of a unitary embedding followed by a partial trace (the Stinespring dilation theorem): for any CPT map  $\mathcal{N}$  there exists an ancillary space of Hermitian matrices  $H_E$  such that

$$\mathcal{N}(\rho) = \text{Tr}_E [U(\rho \otimes |0\rangle_E \langle 0|)U^\dagger]$$

where  $\rho \in H_{d_{\text{in}}}$  and  $U$  is a unitary matrix mapping states  $|\psi\rangle|0\rangle_E$  with  $|\psi\rangle \in H_{d_{\text{in}}}$  to  $H_{d_{\text{out}}} \otimes H_E$ .

The minimum entropy output of a quantum channel  $\mathcal{N}$  is defined by

$$S_{\min}(\mathcal{N}) \equiv \min_{\rho \in H_{d_{\text{in}}, +, 1}} S(\mathcal{N}(\rho)) ,$$

---

\*Electronic address: gour@ucalgary.ca

†Electronic address: friedlan@uic.edu

where  $H_{d_{\text{in}},+,1} \subset H_{d_{\text{in}}}$  is the set of all  $d_{\text{in}} \times d_{\text{in}}$  positive semi-definite matrices with trace one (i.e. density matrices), and  $S(\rho) = -\text{Tr}(\rho \log \rho)$  is the von-Neumann entropy. Since the von-Neumann entropy is concave it follows that the minimization can be taken over all rank one matrices  $\rho = |\psi\rangle\langle\psi|$  in  $H_{d_{\text{in}},+,1}$ .

For any such rank one density matrix  $\rho$  we can define a bipartite pure state  $|\Psi\rangle = U|\psi\rangle|0\rangle_E$  in the bipartite subspace  $\mathcal{K} \equiv \{|\Psi\rangle \mid |\psi\rangle \in H_{d_{\text{in}}}\}$ . We therefore find that the minimum entropy output of the channel  $\mathcal{N}$  can be expressed in terms of the entanglement of the bipartite subspace  $\mathcal{K}$  defined by

$$E(\mathcal{K}) \equiv \min_{|\phi\rangle \in \mathcal{K}, \|\phi\|=1} E(|\phi\rangle),$$

where  $E(|\phi\rangle) \equiv S(\text{Tr}_E(|\phi\rangle\langle\phi|))$  is the entropy of entanglement. In [13] it was pointed out that  $E(\mathcal{K}) = 0$  unless  $\dim \mathcal{K} \leq (d_{\text{out}} - 1)(\dim H_E - 1)$ . This claim follows directly from the fact that the number of (bipartite) states in an unextendible product basis is at least  $d_{\text{out}} + \dim H_E - 1$  [3].

With these notations, the non-additivity of the minimum entropy output of a quantum channel is equivalent to the existence of two subspaces  $\mathcal{K}_1 \subset \mathbb{C}^{n_1} \otimes \mathbb{C}^{m_1}$  and  $\mathcal{K}_2 \subset \mathbb{C}^{n_2} \otimes \mathbb{C}^{m_2}$  such that

$$E(\mathcal{K}_1 \otimes \mathcal{K}_1) < E(\mathcal{K}_1) + E(\mathcal{K}_2).$$

In what follows we will prove the local additivity of entanglement of subspaces, which is equivalent to the local additivity of the minimum entropy output.

The rest of this paper is organized as follows. In section II we find and simplify the first and second directional derivatives of the von-Neumann entropy of entanglement. In section III we prove our main result of local additivity which is stated in Theorem 5 for the non-singular case. In section IV we prove Theorem 5 for the singular case. We end with a discussion in section V.

## II. LOCAL MINIMUM

Let  $\mathcal{K} \subset \mathbb{C}^n \otimes \mathbb{C}^m$  be a subspace of bipartite entangled states. Since the bipartite Hilbert space  $\mathbb{C}^n \otimes \mathbb{C}^m$  is isomorphic to the Hilbert space of all  $n \times m$  complex matrices  $\mathbb{C}^{n \times m}$ , we can view any bipartite state  $|\psi\rangle^{AB} = \sum_{i,j} x_{ij} |i\rangle |j\rangle$  in  $\mathcal{K}$  as an  $n \times m$  matrix  $x$ . The reduced density matrix of  $|\psi\rangle^{AB}$  is then given by  $\rho_r \equiv \text{Tr}_B |\psi\rangle^{AB} \langle\psi| = xx^*$ , and the entropy of entanglement of  $|\psi\rangle^{AB}$  is given by

$$E(x) \equiv -\text{Tr}(xx^* \log xx^*) . \quad (1)$$

In our notations, instead of using a dagger, we use  $x^*$  to denote the hermitian conjugate of the matrix  $x$ .

If  $x \in \mathcal{K}$  is a local minimum of  $E$  in  $\mathcal{K}$ , then there exists a neighbourhood of  $x$  in  $\mathcal{K}$  such that  $x$  is the minimum in that neighbourhood. Any state in the neighbourhood of  $x$  can be written as  $ax + by$ , where  $a, b \in \mathbb{C}$  and  $y \in \mathcal{K}$  is an orthogonal matrix to  $x$ ; i.e.  $\text{Tr}(xy^*) = 0$ . We also assume that the state is normalized so that  $|a|^2 + |b|^2 = 1$ . Now, since the function  $E(x)$  is independent on global phase, we can assume that  $a$  is a positive *real* number. We can also assume that  $b$  is real since we can absorb its phase into  $y$  (adding a phase to  $y$  will not change its orthogonality to  $x$ ). Thus, any normalized state in the neighbourhood of  $x$  can be written as

$$\frac{x + ty}{\sqrt{1 + t^2}} \quad \text{with} \quad \text{Tr}(xy^*) = 0 ,$$

where  $t \equiv b/a$  is a small real number and  $y$  is normalized (i.e.  $\text{Tr}(yy^*) = 1$ ).

### Definition 1.

(a) A matrix  $x \in \mathcal{K}$  is said to be a critical point of  $E(x)$  in  $\mathcal{K}$  if

$$D_y E(x) \equiv \frac{d}{dt} E\left(\frac{x + ty}{\sqrt{1 + t^2}}\right) \Big|_{t=0} = 0 \quad \forall y \in x^\perp$$

where the notation  $D_y E(x)$  indicate that we are taking the directional derivative of  $E$  in the direction of  $y$ , and  $x^\perp \subset \mathcal{K}$  denotes the subspace of all the matrices  $y$  in  $\mathcal{K}$  for which  $\text{Tr}(xy^*) = 0$ .

(b) A matrix  $x \in \mathcal{K}$  is said to be a non-degenerate local minimum of  $E(x)$  in  $\mathcal{K}$  if it is critical and

$$D_y^2 E(x) \equiv \frac{d^2}{dt^2} E\left(\frac{x + ty}{\sqrt{1 + t^2}}\right) \Big|_{t=0} > 0 \quad \forall y \in x^\perp,$$

where we also allow  $D_y^2 E(x) = +\infty$ . Moreover, a critical  $x \in \mathcal{K}$  is said to be degenerate if there exists at least one direction  $y$  such that  $D_y^2 E(x) = 0$ .

In order to prove local additivity we will need to calculate the above directional derivatives. This can be done by expressing the logarithm as an integral [28] (see also [22, 23]). However, in this technique all the quantities are expressed by integrals, and some of these integral expressions do not lead to additivity in a transparent way, as the divided difference method does. We therefore apply below a new technique that is based on the *divided difference* [16, (6.1.17)]. One of the advantages of the divided difference approach, is that it enables one to calculate and express all directional derivatives *explicitly* with no integrals involved. Before introducing the divided difference approach, we will first discuss briefly the affine parametrization.

In our calculations we will assume that  $x$  is diagonal (or equivalently, the bipartite state  $x$  represents is given in its Schmidt form). This assumption follows from the singular value decomposition theorem; namely, we can always find unitary matrices  $u \in \mathbb{C}^{n \times n}$  and  $v \in \mathbb{C}^{m \times m}$  such that  $uxv$  is an  $n \times m$  diagonal matrix with non-negative real numbers (the singular values of  $x$ ) on the diagonal. Since  $E(x) = E(uxv)$  we can assume without loss of generality that  $x$  is a diagonal matrix.

### A. The Affine Parametrization

Up to second order in  $t$  we have

$$\begin{aligned} \rho(t) &\equiv \frac{(x + ty)(x^* + ty^*)}{1 + t^2} = (xx^* + t(xy^* + yx^*) + t^2 yy^*) (1 - t^2) \\ &= xx^* + t(xy^* + yx^*) + t^2(yy^* - xx^*) = \rho + t\gamma_0 + t^2\gamma_1, \end{aligned} \quad (2)$$

where  $\rho = xx^*$ ,  $\gamma_0 \equiv xy^* + yx^*$ , and  $\gamma_1 \equiv yy^* - xx^*$ . Note that  $\text{Tr } \rho = 1$  and  $\text{Tr } \gamma_0 = \text{Tr } \gamma_1 = 0$ , where without loss of generality we assumed  $\text{Tr}(yy^*) = 1$  since we can absorb the normalization factor of  $y$  into  $t$ . We are interested in taking the first and second derivative of

$$E\left(\frac{x + ty}{\sqrt{1 + t^2}}\right) = S(\rho(t)) = S(\rho + t\gamma_0 + t^2\gamma_1).$$

In this section we assume that  $\rho = xx^*$  is an  $n \times n$  non-singular matrix. Denote

$$\sigma(t) \equiv \rho + t\gamma_0.$$

In the next proposition we relate  $S(\rho(t))$  with  $S(\sigma(t))$ .

**Proposition 1.** *Let  $\rho(t), \sigma(t), \rho, \gamma_0$  and  $\gamma_1$  as above. Then*

$$S(\rho(t)) = S(\sigma(t)) - t^2 \text{Tr} [\gamma_1 \log \rho] + O(t^3) \quad (3)$$

*Proof.* Since  $\rho$  is non-singular, also  $\rho(t)$  and  $\sigma(t)$  are non-singular for small enough  $t$ . Thus,  $I - \rho(t) < I$  for small  $t$ . Using the Taylor expansion

$$\log \rho(t) = \log[I - (I - \rho(t))] = - \sum_{n=1}^{\infty} \frac{(I - \sigma(t) - t^2\gamma_1)^n}{n},$$

we get

$$-\text{Tr} [\rho \log \rho(t)] = \sum_{n=1}^{\infty} \frac{1}{n} \text{Tr} [\rho (I - \sigma(t) - t^2\gamma_1)^n].$$

Expanding the term in the trace above up to second order in  $t$  gives

$$\text{Tr} [\rho (I - \sigma(t) - t^2\gamma_1)^n] = \text{Tr} [\rho (I - \sigma(t))^n] + t^2 n \text{Tr} [\rho (I - \rho)^{n-1} \gamma_1] + O(t^3).$$

We therefore have

$$-\text{Tr} [\rho \log \rho(t)] = -\text{Tr} [\rho \log \sigma(t)] + t^2 \sum_{n=1}^{\infty} \text{Tr} [\rho (I - \rho)^{n-1} \gamma_1] + O(t^3).$$

Since  $\rho^{-1} = \sum_{n=1}^{\infty} (I - \rho)^{n-1}$  and  $\text{Tr}(\gamma_1) = 0$  we conclude

$$\text{Tr} [\rho \log \rho(t)] = \text{Tr} [\rho \log \sigma(t)] + O(t^3).$$

Thus,

$$\text{Tr} [\rho(t) \log \rho(t)] = \text{Tr} [\sigma(t) \log \sigma(t)] + t^2 \text{Tr} [\gamma_1 \log \rho] + O(t^3).$$

This completes the proof.  $\square$

This simple relation between  $S(\rho(t))$  and  $S(\sigma(t))$  is very useful since now we can focus on the Taylor expansion of the simpler function  $S(\sigma(t))$ .

## B. The method of divided difference

To calculate the first and second derivatives of  $S(\sigma(t))$ , we first evaluate the Taylor expansion of a complex valued function  $f : \mathbb{C} \rightarrow \mathbb{C}$ , which we later assume can be extended to act on  $n \times n$  complex matrices.

We will make use of the notion of the *divided difference* for  $f$ , which we refer the reader to [16, (6.1.17)] for more details. The divided difference for a function  $f : \mathbb{C} \rightarrow \mathbb{C}$ , given a sequence of distinct complex points,  $\alpha_i \in \mathbb{C}, i = 1, \dots, n$ , is defined for  $i = 0, 1$  by

$$\Delta^0 f(\alpha_1) := f(\alpha_1) \tag{4}$$

$$\Delta^1 f(\alpha_1, \alpha_2) \equiv \Delta f(\alpha_1, \alpha_2) := \frac{f(\alpha_1) - f(\alpha_2)}{\alpha_1 - \alpha_2}, \tag{5}$$

and defined inductively by

$$\Delta^i f(\alpha_1, \dots, \alpha_i, \alpha_{i+1}) = \frac{\Delta^{i-1} f(\alpha_1, \dots, \alpha_{i-1}, \alpha_i) - \Delta^{i-1} f(\alpha_1, \dots, \alpha_{i-1}, \alpha_{i+1})}{\alpha_i - \alpha_{i+1}}, \tag{6}$$

for  $i = 2, 3, \dots, n$ . It is well known that  $\Delta^i f(\alpha_1, \dots, \alpha_i, \alpha_{i+1})$  is a symmetric function in  $\alpha_1, \dots, \alpha_{i+1}$ , e.g. [16, p'393]. For points that are not distinct it is defined by an appropriate limit. For example, for  $x \neq y$  we have

$$\begin{aligned} \Delta f(x, x) &= f'(x) \\ \Delta^2 f(x, x, y) &= \frac{f'(x)}{(x-y)} - \frac{f(x) - f(y)}{(x-y)^2} \end{aligned} \tag{7}$$

$$\Delta^2 f(x, x, x) = \frac{1}{2} f''(x). \tag{8}$$

Note that (8) can be obtained from (7) by setting  $h \equiv y - x \rightarrow 0$  and expanding  $f(y) = f(x + h) = f(x) + hf'(x) + \frac{1}{2}h^2 f''(x) + O(h^3)$ .

**Theorem 2.** Let  $A = \text{diag}(\alpha_1, \dots, \alpha_n) \in \mathbb{C}^{n \times n}$  be a diagonal square matrix, and  $B = [b_{ij}] \in \mathbb{C}^{n \times n}$  be a complex square matrix. Assume that  $f(x) : \mathbb{C} \rightarrow \mathbb{C}$  satisfy one of the following conditions:

1.  $f(x)$  is an analytic function in some domain  $\mathcal{D} \subset \mathbb{C}$  which contains  $\alpha_1, \dots, \alpha_n$ , and can be approximated uniformly in  $\mathcal{D}$  by polynomials.
2.  $\alpha_1, \dots, \alpha_n$  are in a real open interval  $(a, b)$  and  $f$  has two continuous derivatives in  $(a, b)$ .

Then

$$f(A + tB) = f(A) + tL_A(B) + t^2 Q_A(B) + O(t^3) \tag{9}$$

Here  $L_A : \mathbb{C}^{n \times n} \rightarrow \mathbb{C}^{n \times n}$  is a linear operator, and  $Q_B : \mathbb{C}^{n \times n} \rightarrow \mathbb{C}^{n \times n}$  is a quadratic homogeneous noncommutative polynomial in  $B$ . For  $i, j = 1, \dots, n$  we have

$$[L_A(B)]_{ij} = \Delta f(\alpha_i, \alpha_j) b_{ij} = \frac{f(\alpha_i) - f(\alpha_j)}{\alpha_i - \alpha_j} b_{ij} \tag{10}$$

$$[Q_A(B)]_{ij} = \sum_{k=1}^n \Delta^2 f(\alpha_i, \alpha_k, \alpha_j) b_{ik} b_{kj}. \tag{11}$$

In particular

$$\text{Tr}(L_A(B)) = \sum_{j=1}^n f'(\alpha_j) b_{jj} \quad (12)$$

$$\text{Tr}(Q_A(B)) = \sum_{i,j=1}^n \frac{f'(\alpha_i) - f'(\alpha_j)}{2(\alpha_i - \alpha_j)} b_{ij} b_{ji}. \quad (13)$$

*Remark.* The expansion above can be naturally generalized to higher than the second order, but for the purpose of this article, we will only need to expand  $f(A + tB)$  up to the second order in  $t$ . Moreover, for our purposes we will only need to assume that the  $\alpha_i$  are real and the condition 2 on  $f$  holds. We kept condition 1 on  $f$  in the theorem just to be a bit more general.

Note that in all the expressions above, one must identify  $\alpha_i = \alpha_j$  with the limit  $\alpha_j \rightarrow \alpha_i$ . For example, the term

$$\frac{f'(\alpha_i) - f'(\alpha_j)}{2(\alpha_i - \alpha_j)} = \frac{1}{2} f''(\alpha_i) \quad \text{for } \alpha_i = \alpha_j.$$

In particular, note that if  $B$  is diagonal, Eq. (13) gives the known second order term of the Taylor expansion.

*Proof.* From the conditions on  $f$ , it is enough to prove the theorem assuming  $f$  is a polynomial. By linearity, it is enough to prove all the claims for  $f(x) = x^m$ . Clearly, in the expansion

$$(A + tB)^m = A^m + tL_A(B) + t^2Q_A(B) + O(t^3)$$

we must have

$$L_A(B) = \sum_{0 \leq p, q, p+q=m-1} A^p B A^q, \quad (14)$$

$$Q_A(B) = \sum_{0 \leq p, q, r, p+q+r=m-2} A^p B A^q B A^r, \quad (15)$$

where we expanded  $(A + tB)^m$  up to first and second order in  $t$ . All that is left to show is that these matrices coincide with the ones defined in Eqs. (10,11).

Indeed, since  $A$  is diagonal, the matrix elements of the  $L_A(B)$  in Eq.(14) are given by

$$[L_A(B)]_{ij} = \sum_{0 \leq p, q, p+q=m-1} \alpha_i^p \alpha_j^q b_{ij} = \frac{\alpha_i^m - \alpha_j^m}{\alpha_i - \alpha_j} b_{ij},$$

which is equal to the exact same matrix elements given in Eq.(10).

In the same way, since  $A$  is diagonal, observe that the matrix elements of the  $Q_A(B)$  in Eq.(15) are given by

$$[Q_A(B)]_{ij} = \sum_{k=1}^n \sum_{0 \leq p, q, r, p+q+r=m-2} \alpha_i^p \alpha_k^q \alpha_j^r b_{ik} b_{kj}.$$

On the other hand, a straightforward calculation gives for  $f(x) = x^m$

$$\Delta^2 x^m(\alpha_i, \alpha_k, \alpha_j) = \sum_{0 \leq p, q, r, p+q+r=m-2} \alpha_i^p \alpha_k^q \alpha_j^r.$$

Thus, the expressions in Eq. (11) and Eq. (15) for  $Q_A(B)$  are the same.

We now prove Eq. (13). Observe first that Eq. (11) yields

$$\text{Tr}(Q_A(B)) = \sum_{i,j=1}^n \Delta^2 f(\alpha_i, \alpha_i, \alpha_j) b_{ij} b_{ji}, \quad (16)$$

where we have used the symmetry  $\Delta^2 f(\alpha_i, \alpha_j, \alpha_i) = \Delta^2 f(\alpha_i, \alpha_i, \alpha_j)$ . Now, since  $b_{ij} b_{ji}$  is symmetric under an exchange between  $i$  and  $j$ , we can replace  $\Delta^2 f(\alpha_i, \alpha_i, \alpha_j)$  in Eq. (16) with

$$\frac{1}{2} [\Delta^2 f(\alpha_i, \alpha_i, \alpha_j) + \Delta^2 f(\alpha_j, \alpha_j, \alpha_i)] = \frac{1}{2} \Delta f'(\alpha_i, \alpha_j),$$

where for the last equality we used Eq. (7). This completes the proof.  $\square$

We now use the above theorem for the Taylor expansion of the function  $S(\sigma(t))$  in the neighbourhood of  $t = 0$ .

### C. The first and second derivatives of $E(x)$

We first assume that  $\rho$  is non singular. The case where  $\rho$  is singular will be treated separately in section IV.

**Theorem 3.** *Let  $\rho = \text{diag}\{p_1, \dots, p_n\}$  with  $p_j > 0$  for  $j = 1, \dots, n$ . For this case, we get the following expressions:*

$$\begin{aligned} D_y^1 E(x) &\equiv \frac{d}{dt} S(\rho(t)) \Big|_{t=0} = -\text{Tr}(\gamma_0 \log \rho) \\ D_y^2 E(x) &\equiv \frac{d^2}{dt^2} S(\rho(t)) \Big|_{t=0} = -2 \left( \text{Tr}[\gamma_1 \log \rho] + \sum_{j,k} \frac{\log p_j - \log p_k}{2(p_j - p_k)} |(\gamma_0)_{jk}|^2 \right). \end{aligned} \quad (17)$$

*Remark.* The condition for  $x \in \mathcal{K}$  to be critical is  $D_y^1 E(x) = 0$  which is equivalent to  $\text{Tr}[(xy^* + yx^*) \log xx^*] = 0$  for all  $y \in \mathcal{K}$  such that  $\text{Tr}(xy^*) = 0$ . Moreover, if  $x$  is critical then we also have  $D_{iy}^1 E(x) = 0$  for all  $y \in x^\perp \subset \mathcal{K}$ . Hence, if  $x$  is critical we must have

$$\text{Tr}(xy^* \log xx^*) = 0 \quad (18)$$

for all  $y \in x^\perp \subset \mathcal{K}$ .

*Proof.* Theorem 2 implies that

$$S(\rho + t\gamma_0) = S(\rho) + tL_\rho(\gamma_0) + t^2Q_\rho(\gamma_0) + \mathcal{O}(t^3).$$

where  $L_\rho$  and  $Q_\rho$  are the following linear and quadratic forms

$$\begin{aligned} L_\rho(\gamma) &\equiv \sum_{i=1}^n g'(p_i)(\gamma_0)_{ii} \\ Q_\rho(\gamma) &\equiv \sum_{i=1}^n \sum_{j=1}^n \frac{g'(p_i) - g'(p_j)}{2(p_i - p_j)} (\gamma_0)_{ij} (\gamma_0)_{ji}, \end{aligned}$$

and  $g(t) \equiv -t \log t$ . Note that the expressions for  $L_\rho(\gamma_0)$  and  $Q_\rho(\gamma_0)$  above are the traces of the analogous expressions given in theorem 2, since  $S(\rho)$  is defined as the trace of the matrix  $g(\rho) = -\rho \log \rho$ .

Since  $\gamma_0$  is hermitian with zero trace, and  $g'(t) = -1 - \log t$ , we get

$$\begin{aligned} L_\rho(\gamma_0) &= -\text{Tr}(\gamma_0 \log \rho) \\ Q_\rho(\gamma_0) &= -\sum_{j,k} \frac{\log p_j - \log p_k}{2(p_j - p_k)} |(\gamma_0)_{jk}|^2. \end{aligned} \quad (19)$$

Combining this with proposition 1 proves the theorem.  $\square$

In the following lemma, we rewrite the expression in Eq. (17), which will be useful for the proof of local additivity.

**Lemma 4.** *Denote  $w = (y + y^*)/2$ , and  $z = i(y - y^*)/2$ . Denote also  $r_{jk} = \sqrt{p_j/p_k}$ , where  $\{p_j\}_{j=1}^n$  are the eigenvalues of  $\rho = xx^*$ . Then, the expression in Eq.(17) for  $D_y^2 E(x)$  can be rewritten as*

$$D_y^2 E(x) = -2E(x) - \text{Tr}[(yy^* + y^*y) \log \rho] - 2 \sum_{j,k} (|w_{jk}|^2 \Phi(r_{jk}) + |z_{jk}|^2 \Phi(-r_{jk})), \quad (20)$$

where

$$\Phi(r) \equiv \frac{1}{2} \frac{r+1}{r-1} \log r^2, \quad r \in \mathbb{R}, \quad (21)$$

with the identification  $\Phi(1) = 2$ .

*Proof.* The expression in Eq.(17) for  $D_y^2 E(x)$  involves the terms  $|(\gamma_0)_{jk}|^2$ . The matrix  $\gamma_0 = xy^* + yx^* = xy^* + yx$  where  $x = \text{diag}\{\sqrt{p_1}, \dots, \sqrt{p_n}\}$ . Note that  $y^* = w + iz$  and  $y = w - iz$ , where  $w$  and  $z$  are the Hermitian matrices defined in the lemma. Thus,

$$\gamma_0 = xw + wx + i(xz - zx) .$$

In terms of the matrix elements  $w_{jk}$  and  $z_{jk}$  of  $w$  and  $z$ , we have

$$(\gamma_0)_{jk} = (\sqrt{p_j} + \sqrt{p_k})w_{jk} + i(\sqrt{p_j} - \sqrt{p_k})z_{jk} .$$

The square of this expression can be written as

$$|(\gamma_0)_{jk}|^2 = (\sqrt{p_j} + \sqrt{p_k})^2 |w_{jk}|^2 + (\sqrt{p_j} - \sqrt{p_k})^2 |z_{jk}|^2 + i(p_j - p_k)(w_{jk}^* z_{jk} - w_{jk} z_{jk}^*)$$

Moreover, expressing back  $w$  and  $z$  in terms of  $y$  gives  $i(w_{jk}^* z_{jk} - w_{jk} z_{jk}^*) = (|y_{kj}|^2 - |y_{jk}|^2)/2$ . We can therefore write

$$|(\gamma_0)_{jk}|^2 = (\sqrt{p_j} + \sqrt{p_k})^2 |w_{jk}|^2 + (\sqrt{p_j} - \sqrt{p_k})^2 |z_{jk}|^2 + \frac{1}{2}(p_j - p_k)(|y_{kj}|^2 - |y_{jk}|^2) .$$

Substituting this expression, and the value for  $\gamma_1 = yy^* - xx^*$ , into Eq.(17) gives

$$-\frac{1}{2}D_y^2 E(x) = E(x) + \text{Tr} [yy^* \log \rho] + \sum_{j,k} \log \left( \frac{p_j}{p_k} \right) \left\{ \frac{(\sqrt{p_j} + \sqrt{p_k})^2}{2(p_j - p_k)} |w_{jk}|^2 + \frac{(\sqrt{p_j} - \sqrt{p_k})^2}{2(p_j - p_k)} |z_{jk}|^2 + \frac{1}{4}(|y_{kj}|^2 - |y_{jk}|^2) \right\} .$$

Note first that the term

$$\frac{1}{4} \sum_{j,k} \log \left( \frac{p_j}{p_k} \right) (|y_{kj}|^2 - |y_{jk}|^2) = \frac{1}{2} \text{Tr} [(y^* y - yy^*) \log \rho] .$$

Moreover, denoting  $r_{jk} = \sqrt{p_j/p_k}$  we get

$$\frac{(\sqrt{p_j} + \sqrt{p_k})^2}{2(p_j - p_k)} \log \left( \frac{p_j}{p_k} \right) = \frac{(r_{jk} + 1)^2}{2(r_{jk}^2 - 1)} \log r_{jk}^2 = \frac{1}{2} \frac{r_{jk} + 1}{r_{jk} - 1} \log r_{jk}^2 \equiv \Phi(r_{jk})$$

Similarly,

$$\frac{(\sqrt{p_j} - \sqrt{p_k})^2}{2(p_j - p_k)} \log \left( \frac{p_j}{p_k} \right) = \frac{1}{2} \frac{r_{jk} - 1}{r_{jk} + 1} \log r_{jk}^2 = \Phi(-r_{jk})$$

With these notations we get

$$-\frac{1}{2}D_y^2 E(x) = E(x) + \frac{1}{2} \text{Tr} [(yy^* + y^* y) \log \rho] + \sum_{j,k} (|w_{jk}|^2 \Phi(r_{jk}) + |z_{jk}|^2 \Phi(-r_{jk}))$$

This complete the proof.  $\square$

In the rest of the paper we will use the notations

$$\begin{aligned} M_x(y) &\equiv \sum_{j,k=1}^n (|w_{jk}|^2 \Phi(r_{jk}) + |z_{jk}|^2 \Phi(-r_{jk})) = \text{Tr} [w\Phi_\rho^+(w) + z\Phi_\rho^-(z)] \\ \Gamma_x(y) &\equiv -E(x) - \frac{1}{2} \text{Tr} [(y^* y + yy^*) \log xx^*] . \end{aligned} \quad (22)$$

where  $\Phi_\rho^\pm$  are self-adjoint linear operators defining in terms of the Hadamard product between the input matrix and the matrix with elements  $\Phi(\pm r_{jk})$ . That is,

$$[\Phi_\rho^\pm(w)]_{jk} = \Phi(\pm r_{jk})w_{jk} .$$

With these notations we get that  $D_y^2 E(x) > 0$  if and only if

$$M_x(y) < \Gamma_x(y) . \quad (23)$$

### D. The complex structure and additional necessary condition

If  $D_y^2 E(x) > 0$  for all  $y$  orthogonal to  $x$ , then  $D_{iy}^2 E(x)$  is also positive since  $iy$  is orthogonal to  $x$ . That is,

$$M_x(iy) < \Gamma_x(iy) = \Gamma_x(y). \quad (24)$$

Therefore, we get from Eqs. (23,24) that if  $x$  is a non-degenerate local minimum then

$$\frac{1}{2} (M_x(y) + M_x(iy)) = \sum_{j,k} |y_{jk}|^2 \tilde{\Phi}(r_{jk}) < \Gamma_x(y), \quad (25)$$

where

$$\tilde{\Phi}(r) := \frac{1}{2} (\Phi(r) + \Phi(-r)) = \frac{1}{2} \frac{r^2 + 1}{r^2 - 1} \log r^2, \quad (26)$$

with the identification  $\tilde{\Phi}(\pm 1) = 1$ . Let  $\tilde{\Phi}_\rho$  be a self-adjoint linear operator defining in terms of the Hadamard product between the input matrix and the matrix with components  $\tilde{\Phi}(r_{jk})$ . With this notation the necessary condition given in Eq.(25) can be written as

$$\text{Tr} \left[ y^* \tilde{\Phi}_\rho(y) \right] < \Gamma_x(y). \quad (27)$$

A simple analysis of the function  $\tilde{\Phi}$  shows that  $\tilde{\Phi}(r) \geq 1$  with equality if and only if  $r = \pm 1$ . Thus, Eq.(27) also implies the following necessary condition on a local minimum:

$$1 \leq \Gamma_x(y),$$

which can be written as

$$E(y) - E(x) \geq 1 - \frac{1}{2} [S(yy^* \| xx^*) + S(y^* y \| xx^*)] \quad (28)$$

where

$$S(yy^* \| xx^*) \equiv \text{Tr} (yy^* \log yy^*) - \text{Tr} (yy^* \log xx^*)$$

is the relative entropy. Since  $S(yy^* \| xx^*) \geq 0$  with equality if and only if  $yy^* = xx^*$ , we always have  $S(yy^* \| xx^*) > 0$  for  $\text{Tr}(xy^*) = 0$ . Nevertheless, it is possible that  $\text{Tr}(xy^*) = 0$  and yet  $S(yy^* \| xx^*) \leq 1$ . In such cases Eq. (28) gives  $E(y) \geq E(x)$  which is consistent with the fact that  $x$  is a local min.

### III. LOCAL ADDITIVITY

We now state the main result of this paper.

**Theorem 5.** *Let  $x^{(1)}$  and  $x^{(2)}$  be two non-degenerate local minima of  $E(x)$  in  $\mathcal{K}^{(1)} \subset \mathbb{C}^{n_1 \times m_1}$  and  $\mathcal{K}^{(2)} \subset \mathbb{C}^{n_2 \times m_2}$ , respectively. Then,  $x^{(1)} \otimes x^{(2)}$  is a non-degenerate local minimum of  $E(x)$  in  $\mathcal{K}^{(1)} \otimes \mathcal{K}^{(2)}$ . Moreover, if  $x^{(1)}$  is degenerate local minimum and  $x^{(2)}$  is non-degenerate local minimum, then  $x^{(1)} \otimes x^{(2)}$  is a degenerate local minimum.*

The theorem above implies, in particular, that if  $x^{(1)}$  and  $x^{(2)}$  are critical points of  $E(x)$  in  $\mathcal{K}^{(1)}$  and  $\mathcal{K}^{(2)}$ , respectively, then,  $x^{(1)} \otimes x^{(2)}$  is a critical point of  $E(x)$  in  $\mathcal{K}^{(1)} \otimes \mathcal{K}^{(2)}$ . This fact was observed in [6] (see also [24]), and later was stated in [10]. It follows from the linearity in  $y$  of the condition given in Eq. (18) for critical points. More precisely, if  $x^{(1)}$  and  $x^{(2)}$  are critical points, then  $x^{(1)} \otimes x^{(2)}$  is also critical if (see Eq. (18))

$$0 = \text{Tr} \left[ \left( x^{(1)} \otimes x^{(2)} \right) y^* \log \left( x^{(1)} x^{(1)*} \otimes x^{(2)} x^{(2)*} \right) \right] = \\ \text{Tr} \left[ x^{(1)} y^{(1)*} \log(x^{(1)} x^{(1)*}) \right] + \text{Tr} \left[ x^{(2)} y^{(2)*} \log(x^{(2)} x^{(2)*}) \right]$$

for all  $y \in (x^{(1)} \otimes x^{(2)})^\perp$ , where  $y^{(1)*} \equiv \text{Tr}_2[(I \otimes x^{(2)})y^*]$  and  $y^{(2)*} \equiv \text{Tr}_1[(x^{(1)} \otimes I)y^*]$ . In the equation above we used the additivity of the logarithm function under tensor products. Moreover, since  $y \in (x^{(1)} \otimes x^{(2)})^\perp$ , we also have  $y^{(1)} \in (x^{(1)})^\perp$  and  $y^{(2)} \in (x^{(2)})^\perp$ . Thus, if  $x^{(1)}$  and  $x^{(2)}$  are critical points,  $x^{(1)} \otimes x^{(2)}$  is also critical [29].

In the following subsection we provide one of the main ingredients for the local additivity of the von-Neumann entropy output of a quantum channel.



### A. The Subadditivity of $\Phi_\rho^\pm$

**Lemma 6.** Let  $\Phi, \tilde{\Phi} : \mathbb{R} \rightarrow \mathbb{R}$  be defined as in Eq.(21) and Eq. (26), respectively. Then, for any  $r, s \in \mathbb{R}$  the following holds:

$$\Phi(rs) \leq \tilde{\Phi}(r) + \tilde{\Phi}(s) \quad (29)$$

with equality if and only if  $r = s$ . In the operator language of Eqs. (22,27), the inequality (29) can be expressed as

$$\Phi_{\rho^A \otimes \rho^B}^\pm \leq \tilde{\Phi}_{\rho^A \otimes I^B} + \tilde{\Phi}_{I^A \otimes \rho^B} = \tilde{\Phi}_{\rho^A} \otimes I^B + I^A \otimes \tilde{\Phi}_{\rho^B} . \quad (30)$$

where two operators satisfies  $O_1 \leq O_2$  if and only if  $\text{Tr}[y^* O_1 y] \leq \text{Tr}[y^* O_2 y]$  for all  $y$ .

*Proof.* We need to prove that

$$\frac{rs+1}{rs-1} \log(rs^2) \leq \frac{r^2+1}{r^2-1} \log r^2 + \frac{s^2+1}{s^2-1} \log s^2$$

This inequality is equivalent to

$$\left( \frac{r^2+1}{r^2-1} - \frac{rs+1}{rs-1} \right) \log r^2 + \left( \frac{s^2+1}{s^2-1} - \frac{rs+1}{rs-1} \right) \log s^2 \geq 0$$

which is equivalent to

$$\frac{s-r}{rs-1} (f(r) - f(s)) \geq 0 , \quad (31)$$

where

$$f(r) \equiv \frac{r}{r^2-1} \log r^2 .$$

That is, we need to prove that  $f(r) \geq f(s)$  if  $(s-r)/(rs-1) > 0$  and  $f(r) \leq f(s)$  if  $(s-r)/(rs-1) < 0$ . From symmetry under exchange of  $r$  and  $s$ , both cases are equivalent, and therefore without loss of generality we assume  $(s-r)/(rs-1) > 0$ . This inequality is satisfied if **(a)**  $s > r$  and  $rs > 1$  or **(b)**  $s < r$  and  $rs < 1$ . A simple analysis of the function  $f(r)$  shows that  $f$  is odd, and it is monotonically increasing for  $-1 \leq r \leq 1$  and monotonically decreasing for  $|r| > 1$ . Moreover, note that  $f(1/r) = f(r)$ .

Consider case **(a)**: If  $s > r > 1$  then  $f(r) \geq f(s)$  since  $f$  is monotonically decreasing in this region. In the same way if  $-1 > s > r$  then  $f(r) \geq f(s)$ . Another possibility in this case is that  $0 < r < 1 < 1/r < s$ . But since both  $r$  and  $1/s$  are positive and smaller than 1, we get  $f(r) \geq f(1/s) = f(s)$ , where we have used the fact that  $f(r)$  is monotonically increasing for  $|r| \leq 1$ . The last possibility in this case is that  $1/r > s > -1 > r$ . For this last possibility both  $s$  and  $1/r$  are negative numbers bigger than  $-1$  and in this region  $f$  is monotonically increasing. Thus,  $f(r) = f(1/r) \geq f(s)$ .

Consider case **(b)**: First note that if  $s < 0 < r$  then  $f(s) < 0 < f(r)$ , and if  $-1 < s < r < 1$  then  $f(r) \geq f(s)$  since  $f$  is monotonically increasing in this region. Another possibility in this case is that  $s < 1 < r < 1/s$ . But since both  $r$  and  $1/s$  are positive and bigger than 1, we get  $f(r) \geq f(1/s) = f(s)$ , where we have used the fact that  $f(r)$  is monotonically decreasing for  $r \geq 1$ . Finally, the last possibility in this case is that  $1/r < s < -1 < r$ . For this last possibility both  $s$  and  $1/r$  are negative numbers smaller than  $-1$  and in this region  $f$  is monotonically decreasing. Thus,  $f(r) = f(1/r) \geq f(s)$ .

In order to prove the equality conditions, we need to show that the expression in Eq.(31) equals zero if and only if  $s = r$ . Before proceeding to prove that, we check the case  $r = 1/s$ . In this case,  $\Phi(rs) = \Phi(1) = 2$  and  $\tilde{\Phi}(s) = \tilde{\Phi}(1/r) = \tilde{\Phi}(r)$ . That is, if  $r = 1/s$  then the equality in Eq. (29) holds if and only if  $\tilde{\Phi}(r) = 1$ . As pointed out earlier,  $\tilde{\Phi}(r) = 1$  if and only if  $r = \pm 1$ . We therefore conclude that if  $r = 1/s$  then the equality in Eq. (29) holds if and only if  $r = s = \pm 1$ . Assume now  $rs \neq 1$ . In this case, the expression in Eq.(31) equals zero if and only if  $f(r) = f(s)$ . However, a simple analysis of the function  $f(r)$  implies that  $f(r) = f(s)$  if and only if  $r = s$  or  $r = 1/s$ . Since we assumed  $rs \neq 1$ , we get that  $r = s$ . This completes the proof.  $\square$

### B. Proof of Theorem 5

We can assume without loss of generality that  $n_1 = m_1$ ,  $n_2 = m_2$ . This can always be done by adding zero rows or columns. However, in this part of the proof we also assume that both  $x^{(1)}$  and  $x^{(2)}$  are non-singular. The singular case is treated separately in section IV. From the singular valued decomposition (see the argument below definition 1) we can assume without loss of generality that  $x^{(1)} = \text{diag}\{\sqrt{p_1}, \dots, \sqrt{p_{n_1}}\}$  and  $x^{(2)} = \text{diag}\{\sqrt{q_1}, \dots, \sqrt{q_{n_2}}\}$ , where  $p_i$  and  $q_j$  are positive and  $\sum_{i=1}^{n_1} p_i = \sum_{j=1}^{n_2} q_j = 1$ .

We first assume that both  $x^{(1)}$  and  $x^{(2)}$  are non-degenerate local minima. We need to show that  $D_y^2 E(x) > 0$  for all  $y \in x^\perp$ , where  $x \equiv x^{(1)} \otimes x^{(2)}$ . The most general  $y \in (x^{(1)} \otimes x^{(2)})^\perp$  can be written as

$$y = c_1 x^{(1)} \otimes y^{(2)} + c_2 y^{(1)} \otimes x^{(2)} + c_3 y', \quad (32)$$

where  $y^{(1)} \in (x^{(1)})^\perp$ ,  $y^{(2)} \in (x^{(2)})^\perp$ , and  $y' \in (x^{(1)})^\perp \otimes (x^{(2)})^\perp$  are all normalized. The numbers  $c_j$  can be chosen to be real because we can absorb their phases in  $y^{(1)}$ ,  $y^{(2)}$ , and  $y'$ . They also satisfy  $c_1^2 + c_2^2 + c_3^2 = 1$ , so that  $y$  is normalized.

Consider first the simple case where  $y = x^{(1)} \otimes y^{(2)}$ . In this case,

$$\begin{aligned} E\left(\frac{x + ty}{\sqrt{1+t^2}}\right) &= E\left(x^{(1)} \otimes \frac{x^{(2)} + ty^{(2)}}{\sqrt{1+t^2}}\right) \\ &= E\left(x^{(1)}\right) + E\left(\frac{x^{(2)} + ty^{(2)}}{\sqrt{1+t^2}}\right). \end{aligned} \quad (33)$$

Since  $x^{(2)}$  is a non-degenerate local minimum, we must have  $D_y^2 E(x) > 0$ . The case  $y = y^{(1)} \otimes x^{(2)}$  is similar.

Consider now the case in which  $y \in (x^{(1)})^\perp \otimes (x^{(2)})^\perp$ . Using its Schmidt decomposition, we can write it as

$$y = \sum_l c_l y_l^{(1)} \otimes y_l^{(2)}, \quad (34)$$

where

$$\text{Tr}[y_l^{(1)} y_{l'}^{(1)*}] = \text{Tr}[y_l^{(2)} y_{l'}^{(2)*}] = \delta_{ll'}, \quad (35)$$

and  $c_l$  are real numbers such that  $\sum_l c_l^2 = 1$ .

By definition we have

$$M_x(y) = \text{Tr} \left[ w^{AB} \Phi_{\rho^A \otimes \rho^B}^+(w^{AB}) + z^{AB} \Phi_{\rho^A \otimes \rho^B}^-(z^{AB}) \right], \quad (36)$$

where  $w^{AB} = (y^* + y)/2$ ,  $z^{AB} = i(y^* - y)/2$ ,  $\rho^A \equiv x^{(1)} x^{(1)*}$  and  $\rho^B \equiv x^{(2)} x^{(2)*}$ .

Applying lemma 6 both to  $\Phi_{\rho^A \otimes \rho^B}^\pm$  gives:

$$\begin{aligned} M_x(y) &\leq \text{Tr} \left[ w^{AB} \tilde{\Phi}_{\rho^A \otimes I^B}(w^{AB}) + w^{AB} \tilde{\Phi}_{I^A \otimes \rho^B}(w^{AB}) + z^{AB} \tilde{\Phi}_{\rho^A \otimes I^B}(z^{AB}) + z^{AB} \tilde{\Phi}_{I^A \otimes \rho^B}(z^{AB}) \right] \\ &= \text{Tr} \left[ y^* \tilde{\Phi}_{\rho^A \otimes I^B}(y) + y^* \tilde{\Phi}_{I^A \otimes \rho^B}(y) \right]. \end{aligned}$$

where  $I^A$  and  $I^B$  are the identity matrices in the respective spaces, and in the last equality we have used the definitions  $w^{AB} = (y^* + y)/2$  and  $z^{AB} = i(y^* - y)/2$ . Now, but substituting (34) into the above equation we get

$$M_x(y) \leq \sum_l c_l^2 \text{Tr} \left[ y_l^{(1)*} \tilde{\Phi}_{\rho^A}(y_l^{(1)}) + y_l^{(2)*} \tilde{\Phi}_{\rho^B}(y_l^{(2)}) \right],$$

where we have used the orthogonality relations in Eq. (35). Combining this with Eq. (27) gives

$$M_x(y) < \sum_l c_l^2 \left( \Gamma_{x^{(1)}}(y_l^{(1)}) + \Gamma_{x^{(2)}}(y_l^{(2)}) \right) = \Gamma_x(y) \quad (37)$$

where the last equality can be verified from the orthogonality relations given in Eq. (35), and the fact that

$$\log x x^* = \log x^{(1)} x^{(1)*} \otimes I^B + I^A \otimes \log x^{(2)} x^{(2)*}. \quad (38)$$

This completes the proof for  $y \in (x^{(1)})^\perp \otimes (x^{(2)})^\perp$ .

Consider now the most general case where  $y \in x^\perp$  has the form given in Eq. (32). Denote

$$w^{AB} = \frac{1}{2}(y^* + y) = c_1 x^{(1)} \otimes w^{(2)} + c_2 w^{(1)} \otimes x^{(2)} + c_3 w', \quad (39)$$

where  $w' = (y'^* + y')/2$  and we have used

$$\begin{aligned} \frac{1}{2} \left( x^{(1)*} \otimes y^{(2)*} + x^{(1)} \otimes y^{(2)} \right) &= x^{(1)} \otimes \frac{1}{2} \left( y^{(2)*} + y^{(2)} \right) \equiv x^{(1)} \otimes w^{(2)} \\ \frac{1}{2} \left( y^{(1)*} \otimes x^{(2)*} + y^{(1)} \otimes x^{(2)} \right) &= \frac{1}{2} \left( y^{(1)*} + y^{(1)} \right) \otimes x^{(2)} \equiv w^{(1)} \otimes x^{(2)}. \end{aligned}$$

In the above equation we used the fact that  $x^{(1)}$  and  $x^{(2)}$  are square diagonal matrices with their singular values on the diagonal. We would like to substitute the expression in Eq. (39) for  $w^{AB}$ , into the expression for  $M_x(y)$  given in Eq. (36). By doing that we will get expressions with several cross terms. We argue that these cross terms vanish. To see that consider for example the cross term

$$c_1 c_3 \text{Tr} \left[ x^{(1)} \otimes w^{(2)} \Phi_{\rho^A \otimes \rho^B}^+(w') \right],$$

and recall that  $\rho^A \equiv x^{(1)} x^{(1)*}$  and  $\rho^B \equiv x^{(2)} x^{(2)*}$ . Since  $\Phi_{\rho^A \otimes \rho^B}^+$  is self-adjoint, the above expression can be written as

$$c_1 c_3 \text{Tr} \left[ x^{(1)} \otimes w^{(2)} \Phi_{\rho^A \otimes \rho^B}^+(w') \right] = c_1 c_3 \text{Tr} \left[ w' \Phi_{\rho^A \otimes \rho^B}^+(x^{(1)} \otimes w^{(2)}) \right] = c_1 c_3 \text{Tr} \left[ w' \left( x^{(1)} \otimes \Phi_{\rho^B}^+(w^{(2)}) \right) \right]$$

where in the last equality we used the identity  $\Phi_{\rho^A \otimes \rho^B}^+(x^{(1)} \otimes w^{(2)}) = x^{(1)} \otimes \Phi_{\rho^B}^+(w^{(2)})$ . This identity follows from the definition of  $\Phi_\rho^+$ , when working with a basis in which  $x^{(1)}$  is diagonal. Now, since the partial trace  $\text{Tr}_1[w'(x^{(1)} \otimes B)] = 0$  for all matrices  $B$ , we have

$$c_1 c_3 \text{Tr} \left[ x^{(1)} \otimes w^{(2)} \Phi_{\rho^A \otimes \rho^B}^+(w') \right] = 0.$$

In the same way, we see that all the other cross terms vanish. Moreover, denote

$$z^{AB} = \frac{i}{2}(y^* - y) = c_1 x^{(1)} \otimes z^{(2)} + c_2 z^{(1)} \otimes x^{(2)} + c_3 z',$$

where  $z^{(1)}$ ,  $z^{(2)}$ , and  $z'$  are defined similarly to  $w^{(1)}$ ,  $w^{(2)}$ , and  $w'$ . Substituting this expression for  $z^{AB}$  in Eq. (36) will also lead to vanishing cross terms. To summarize, by substituting the above expressions for  $z^{AB}$  and  $w^{AB}$  in Eq. (36) we get

$$M_x(y) = c_1^2 M_x(x^{(1)} \otimes y^{(2)}) + c_2^2 M_x(y^{(1)} \otimes x^{(2)}) + c_3^2 M_x(y')$$

However, since we already proved that  $x$  is a non-degenerate local minimum in the directions  $x^{(1)} \otimes y^{(2)}$ ,  $y^{(1)} \otimes x^{(2)}$ , and  $y'$ , we get

$$M_x(y) < c_1^2 \Gamma_x(x^{(1)} \otimes y^{(2)}) + c_2^2 \Gamma_x(y^{(1)} \otimes x^{(2)}) + c_3^2 \Gamma_x(y') \quad (40)$$

Now, note the orthogonality relations in the partial traces:  $\text{Tr}_1[(x^{(1)} \otimes y^{(2)})(y')^*] = \text{Tr}_2[(x^{(1)} \otimes y^{(2)})(y')^*] = 0$  and  $\text{Tr}_1[(y^{(1)} \otimes x^{(2)})(y')^*] = \text{Tr}_2[(y^{(1)} \otimes x^{(2)})(y')^*] = 0$ . With these relations and from Eq. (38) we get that the expression in the RHS of Eq.(40) is equal to  $\Gamma_x(y)$ . This completes the proof of the main part of the theorem.

To prove the second part of the theorem, assume that  $x^{(1)}$  is degenerate local minimum and  $x^{(2)}$  is a non-degenerate local minimum. Following the exact same lines of the proof above we get that  $M_x(y') < \Gamma_x(y')$  for  $y' \in (x^{(1)})^\perp \otimes (x^{(2)})^\perp$ . This is clear from Eq. (37) and the one above it, where we use the fact that

$$\text{Tr} \left[ y_l^{(2)*} \tilde{\Phi}_{\rho^B}(y_l^{(2)}) \right] < \Gamma_{x^{(2)}}(y_l^{(2)})$$

since  $x^{(2)}$  is a non-degenerate local minimum. Similarly, if  $y = x^{(1)} \otimes y^{(2)}$  we get  $M_x(y) < \Gamma_x(y)$ . The only  $y \in x^\perp$  for which it is possible to have  $M_x(y) = \Gamma_x(y)$  is  $y = y^{(1)} \otimes x^{(2)}$ . However, in this case

$$E \left( \frac{x + ty}{\sqrt{1+t^2}} \right) = E \left( \frac{x^{(1)} + ty^{(1)}}{\sqrt{1+t^2}} \right) + E \left( x^{(2)} \right), \quad (41)$$

so  $x$  is a local minimum in this direction as well. Hence,  $x$  is a degenerate local minimum. This completes the proof of the second part of the theorem.

#### IV. THE SINGULAR CASE

In the previous section, we were able to derive the first and second directional derivatives  $D_y^1 E(x)$  and  $D_y^2 E(x)$  assuming  $x$  is non-singular. In this section we consider the case where  $x$  is singular. While the expression for  $D_y^1 E(x)$  is the same as in the previous section, the expression for the second derivative is not the same for the singular case. In particular, in the singular case it is possible that  $D_y^2 E(x) \equiv \frac{d^2}{dt^2} S(\rho(t))|_{t=0}$  diverge. Nevertheless, we will see in this section that even if  $x$  is singular,  $E(x)$  is additive.

For simplicity of the exposition, we will consider here subspaces  $\mathcal{K} \subset \mathbb{C}^n \otimes \mathbb{C}^m$ , where  $n = m$ , since we can always embed  $\mathcal{K}$  in  $\mathbb{C}^{\max\{n,m\}} \otimes \mathbb{C}^{\max\{n,m\}}$ . The following theorem provides the criterion for the divergence of the second derivative.

**Theorem 7.** *Let  $x, y \in \mathcal{K} \subset \mathbb{C}^{n \times n}$ ,  $\text{Tr } xx^* = \text{Tr } yy^* = 1$  and  $\text{Tr } (xy^*) = 0$ . Change the standard orthonormal base in  $\mathbb{C}^n$  to a new orthonormal base such that  $x$  and  $y$  have the forms*

$$x = \begin{bmatrix} x_{11} & 0_{r,n-r} \\ 0_{n-r,r} & 0_{n-r,n-r} \end{bmatrix} \quad \text{and} \quad y = \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{bmatrix}, \quad (42)$$

where  $r$  is the rank of  $x$ ,  $0_{i,j}$  are  $i \times j$  zero matrices, and  $x_{11}, y_{11} \in \mathbb{C}^{r \times r}$ . Then

$$S(\rho(t)) = f(t) - (K + tg(t))t^2 \log t^2, \quad K = \text{Tr } (y_{22} y_{22}^*), \quad (43)$$

where  $f(t), g(t)$  are analytic functions in a neighbourhood of 0. Hence  $D_y^2 E(x) = +\infty$  if and only if  $y_{22} \neq 0$ . Furthermore, if  $y_{22} = 0$  then either  $g(t) \equiv 0$  or  $g(t) = at^{2k-1}(1 + O(t))$ , where  $a > 0$  and  $k$  is a positive integer.

A much weaker version of the theorem above can be found in [10]. For the clarity of the exposition in this section, we leave the proof of Theorem 7 to appendix A.

From the theorem above it follows that w.l.o.g we can set  $y_{22} = 0$  since otherwise the second derivative is  $+\infty$ . This will be useful when proving local additivity for the singular case. However, in the tensor product space,  $y$  can be written as in Eq. (34). Hence, while we assume that the  $(2,2)$  block of the bipartite state  $y$  is zero, it is not immediately obvious that the  $(2,2)$  blocks of the one-party states  $y_i^{(1)}$  and  $y_i^{(2)}$  are also zero. Nevertheless, this is indeed the case as we show now.

##### A. Tensor product structure in the singular case

Let  $\mathcal{K} \subset \mathbb{C}^{n \times n}$  be a subspace of matrices that are partitioned as in Eq. (42). We assume that  $\mathcal{K}$  contains a matrix

$$x = \begin{bmatrix} x_{11} & 0 \\ 0 & 0 \end{bmatrix}, \quad \text{Tr } (x_{11}^* x_{11}) = 1. \quad (44)$$

We now choose a following orthonormal base  $x_1, \dots, x_p, y_1, \dots, y_q, z_1, \dots, z_r, w_1, \dots, w_s \in \mathcal{K}$ . First,  $x_1 = x$ . Then

1.  $x_1, \dots, x_p$  is an orthonormal basis of the subspace of  $\mathcal{K}$  of matrices of the form  $\begin{bmatrix} * & 0 \\ 0 & 0 \end{bmatrix}$ . (It is possible that  $p = 1$ .)
2.  $x_1, \dots, x_p, y_1, \dots, y_q$  is an orthonormal basis of the subspace of  $\mathcal{K}$  of matrices of the form  $\begin{bmatrix} * & * \\ 0 & 0 \end{bmatrix}$ . (It is possible that  $q = 0$ .)
3.  $x_1, \dots, x_p, y_1, \dots, y_q, z_1, \dots, z_r$  is an orthonormal basis of the subspace of  $\mathcal{K}$  of matrices of the form  $\begin{bmatrix} * & * \\ * & 0 \end{bmatrix}$ . (It is possible that  $r = 0$ .)
4.  $x_1, \dots, x_p, y_1, \dots, y_q, z_1, \dots, z_r, w_1, \dots, w_s$  is an orthonormal basis of  $\mathcal{K}$ . (It is possible that  $s = 0$ .)

We observe the following

1. The projections of  $x_1, \dots, x_p$  on the block  $(1,1)$  are linearly independent.
2. The projections of  $y_1, \dots, y_q$  on the block  $(1,2)$  are linearly independent if  $q \geq 1$ .

3. The projections of  $z_1, \dots, z_r$  on the block  $(2, 1)$  are linearly independent if  $r \geq 1$ .
4. The projections of  $w_1, \dots, w_s$  on the block  $(2, 2)$  are linearly independent if  $s \geq 1$ .

We now consider two subspaces  $\mathcal{K}_i \subset \mathbb{C}^{n_i \times n_i}$  for  $i = 1, 2$ . We consider here the most complicated case in which both matrices  $x^{(1)} \in \mathcal{K}_1$  and  $x^{(2)} \in \mathcal{K}_2$  are singular. So we assume that each  $x^{(i)}$  has the form (44). For  $i = 1, 2$  we form orthonormal bases

$$x_1^{(i)}, \dots, x_{p_i}^{(i)}, y_1^{(i)}, \dots, y_{q_i}^{(i)}, z_1^{(i)}, \dots, z_{r_i}^{(i)}, w_1^{(i)}, \dots, w_{s_i}^{(i)} \in \mathcal{K}_i \quad (45)$$

exactly as above. We now form a tensor product of  $\mathcal{K}_1 \otimes \mathcal{K}_2$  with respect to the partitions of  $\mathcal{K}_1, \mathcal{K}_2$  as above.

Let

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \in \mathcal{K}_1, \quad B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \in \mathcal{K}_2. \quad (46)$$

We then agree that the partition in  $\mathcal{K}_1 \otimes \mathcal{K}_2$  is of the form as the following partition of  $A \otimes B$ :

$$A \otimes B = \begin{bmatrix} A_{11} \otimes B_{11} & A_{11} \otimes B_{12} & A_{12} \otimes B_{11} & A_{12} \otimes B_{12} \\ A_{11} \otimes B_{21} & A_{11} \otimes B_{22} & A_{12} \otimes B_{21} & A_{12} \otimes B_{22} \\ A_{21} \otimes B_{11} & A_{21} \otimes B_{12} & A_{22} \otimes B_{11} & A_{22} \otimes B_{12} \\ A_{21} \otimes B_{21} & A_{21} \otimes B_{22} & A_{22} \otimes B_{21} & A_{22} \otimes B_{22} \end{bmatrix}. \quad (47)$$

**Lemma 8.** *Let  $\mathcal{K}_1, \mathcal{K}_2$  be two subspaces in  $\mathbb{C}^{n_1 \times n_1}$  and  $\mathbb{C}^{n_2 \times n_2}$ , respectively. Let  $C = [C_{ij}]_{i,j=1}^4 \in \mathcal{K}_1 \otimes \mathcal{K}_2$  be partitioned as in (47). Suppose that  $C \neq 0$  and  $C_{ij} = 0$  for  $i, j \geq 2$ . Write  $C$  as a linear combination of the tensor products of the bases of  $\mathcal{K}_1$  and  $\mathcal{K}_2$ , chosen as in (45). Then each term in this linear combination of  $C$  is of the form  $\alpha f \otimes g$ , where  $\alpha \in \mathbb{C}$ ,  $f \in \mathcal{K}_1$ ,  $g \in \mathcal{K}_2$ , and both  $f$  and  $g$  have the form  $\begin{bmatrix} * & * \\ * & 0 \end{bmatrix}$ .*

*Remark.* It is also possible to show that at least one of the matrices  $f$  and  $g$  must have the form  $\begin{bmatrix} * & * \\ 0 & 0 \end{bmatrix}$  or  $\begin{bmatrix} * & 0 \\ * & 0 \end{bmatrix}$ . However, we will not be using it here.

*Proof.* Suppose the expansion of  $C$  contains a term of the form  $w_i^{(1)} \otimes w_j^{(2)}$ . Look at the block  $(4, 4)$ . The contribution of the expansion of  $C$  to this block only comes from the tensor products projections of  $w_i^{(1)}$  and  $w_j^{(2)}$  on the block  $(2, 2)$ . Since all these projections are linearly independent we must have that  $C_{44} \neq 0$  contrary to our assumption.

Assume now that the expansion of  $C$  contains  $w_i^{(1)} \otimes z_j^{(2)}$ . Since the expansion of  $C$  does not have terms  $w_i^{(1)} \otimes w_j^{(2)}$ , the contribution to the block  $C_{43}$  comes only from the projection of  $w_i^{(1)}$  on the block  $(2, 2)$  and the projection of  $z_j^{(2)}$  on the block  $(2, 1)$ . Again as all these projections are linearly independent we deduce that  $C_{43} \neq 0$ , contrary to our assumptions.

Similarly, there are no terms in the expansion of  $C$  of the form  $w_i^{(1)} \otimes y_j^{(2)}$ , since  $C_{34} = 0$ , and there are no terms in the expansion of  $C$  of the form  $w_i^{(1)} \otimes x_j^{(2)}$  since  $C_{33} = 0$ . That is, we have shown that the matrices  $w_i^{(1)}$  do not appear in the expansion of  $C$ . In exactly the same way, there are no terms in the expansion of  $C$  of the form  $z_i^{(1)} \otimes w_j^{(2)}$ ,  $y_i^{(1)} \otimes w_j^{(2)}$ , and  $x_i^{(1)} \otimes w_j^{(2)}$ , since  $C_{42} = 0$ ,  $C_{24} = 0$ , and  $C_{22} = 0$ , respectively. This completes the proof.  $\square$

## B. Local additivity in the singular case

In this subsection we prove Theorem 5 for the case in which  $x^{(1)}$  and  $x^{(2)}$  are *singular* local minima of  $\mathcal{K}^{(1)}$  and  $\mathcal{K}^{(2)}$ , respectively. We therefore choose bases such that  $x^{(1)}$  and  $x^{(2)}$  are of the form given in Eq. (42), and denote by  $r_1$  and  $r_2$  their respective ranks.

Assume first that both  $x^{(1)}$  and  $x^{(2)}$  are *non-degenerate* local minima. We need to show that  $D_y^2 E(x) > 0$  for all  $y \in x^\perp$ , where  $x \equiv x^{(1)} \otimes x^{(2)}$ . Note that the partition of  $x = [x_{ij}]_{i,j=1}^4$  as in Eq. (47) gives  $x_{ij} = 0$  for all  $i, j = 1, 2, 3, 4$  except for  $x_{11} = x_{11}^{(1)} \otimes x_{11}^{(2)}$ .

The most general  $y \in (x^{(1)} \otimes x^{(2)})^\perp$  can be written as in Eq. (32), where  $y'$  is of the form given in Eq. (34). Consider now the partition of  $y = [y_{ij}]_{i,j=1}^4$  as in Eq. (47). From Theorem 7 we know that  $D_y^2 E(x) = +\infty$  unless

$y_{ij} = 0$  for all  $i, j = 2, 3, 4$ . We therefore assume now that  $y_{ij} = 0$  for all  $i, j = 2, 3, 4$ . In this case, Lemma 8 implies that all the matrices  $y_i^{(1)}$  and  $y_i^{(2)}$  in Eq. (34) have the form

$$\begin{bmatrix} * & * \\ * & 0 \end{bmatrix}.$$

That is, their  $(2, 2)$  block is zero. For this reason, we replace each subspace  $\mathcal{K}^{(i)} \subset \mathbb{C}^{n_i \times n_i}$  ( $i = 1, 2$ ) with a smaller subspace  $\mathcal{U}^{(i)} \subset \mathcal{K}^{(i)}$  such that each matrix in the basis of  $\mathcal{U}^{(i)}$  has zeros on the  $(2, 2)$  block. It is left to prove that  $x \equiv x_1 \otimes x_2$  is local minimum in  $\mathcal{U}^{(1)} \otimes \mathcal{U}^{(2)}$ .

Consider the new subspace  $\mathcal{U}_\epsilon^{(i)}$ , for  $\epsilon > 0$ , where in the orthonormal basis of  $\mathcal{U}^{(i)}$ , we change only the first matrix  $x^{(i)}$ , i.e. the local minimum matrix, with the normalized diagonal matrix

$$x_\epsilon^{(i)} \equiv \frac{1}{\sqrt{1 + (n_i - r_i)\epsilon^2}} \begin{bmatrix} x_{11}^{(i)} & 0_{r_i, n_i - r_i} \\ 0_{n_i - r_i, r_i} & \epsilon I_{n_i - r_i} \end{bmatrix} \quad i = 1, 2$$

where  $0_{i,j}$  are  $i \times j$  zero matrices and  $I_{n_i - r_i}$  are  $(n_i - r_i) \times (n_i - r_i)$  identity matrices.

**Lemma 9.** *Assume  $x^{(i)}$  is a non-degenerate local minimum in  $\mathcal{U}^{(i)}$ , then  $x_\epsilon^{(i)}$  is a non-degenerate local minimum in  $\mathcal{U}_\epsilon^{(i)}$ . Moreover, there exists  $\delta > 0$  and  $\epsilon_0 > 0$  such that if  $\epsilon < \epsilon_0$  then  $D_{y^{(i)}}^2 E(x_\epsilon^{(i)}) > \delta$  for all  $y^{(i)} \in (x_\epsilon^{(i)})^\perp$ .*

*Proof.* For simplicity of the exposition we remove the superscript  $(i)$  from  $x^{(i)}$  and denote  $d \equiv n - r$ . That is, consider

$$x = \begin{bmatrix} x_{11} & 0_{r,d} \\ 0_{d,r} & 0_{d,d} \end{bmatrix} \text{ and } x_\epsilon \equiv \frac{1}{\sqrt{1 + d\epsilon^2}} \begin{bmatrix} x_{11} & 0_{r,d} \\ 0_{d,r} & \epsilon I_d \end{bmatrix}.$$

We need to show that if  $x$  is a non-degenerate local minimum in  $\mathcal{U}$  then  $x_\epsilon$  is a non-degenerate local minimum in  $\mathcal{U}_\epsilon$  for small enough  $\epsilon$ .

First, we need to show that  $x_\epsilon$  remains critical. Indeed, since the condition (18) for criticality is satisfied for  $x$ , it is also satisfied for  $x_\epsilon$ . This is because  $x_\epsilon$  is a diagonal matrix and all  $y \in x_\epsilon^\perp \subset \mathcal{U}$  is of the form  $\begin{bmatrix} * & * \\ * & 0_{d,d} \end{bmatrix}$ .

Second, we need to show that  $D_y^2 E(x_\epsilon) > \delta$ . In Appendix B we show that  $D_y^2 E(x_\epsilon)$  does *not* diverge in the limit  $\epsilon \rightarrow 0$  (assuming  $y_{jk} = 0$  when both  $j > r$  and  $k > r$ ). Now, since we assume  $D_y^2 E(x) > 0$  for all  $y \in x^\perp$ , we can also assume that there exist  $\delta' > 0$  such that  $D_y^2 E(x) > \delta'$  for all  $y \in x^\perp$ . This is true because the set of all normalized matrices in  $x^\perp$  is compact. Hence, from the nice behaviour of  $D_y^2 E(x_\epsilon)$  in the limit  $\epsilon \rightarrow 0$  (see Appendix B), we get that for small enough  $\epsilon$  there exists  $\delta > 0$  such that  $D_y^2 E(x_\epsilon) > \delta$  for all  $y \in (x_\epsilon)^\perp$ . This completes the proof of the lemma.  $\square$

We now apply Theorem 5 to the non-singular case of  $x_\epsilon \equiv x_\epsilon^{(1)} \otimes x_\epsilon^{(2)}$ . From Lemma 9, the second derivatives  $D_{y^{(i)}}^2 E(x_\epsilon^{(i)}) > \delta$  for all  $y^{(i)} \in (x^{(i)})^\perp$  and  $i = 1, 2$ . Thus, we get that  $D_y^2 E(x_\epsilon) > 2\delta$  for all  $y \in (x_\epsilon)^\perp$ . We obtain it by following precisely the same steps of the proof of Theorem 5 (in the non-singular case). Letting  $\epsilon \rightarrow 0$  we deduce that in the direction of  $y$  the second derivative at  $x^{(1)} \otimes x^{(2)}$  is strictly positive (greater or equal to  $2\delta$ ). This complete the proof of the main part of theorem 5 for the singular case.

To proof the second part of the theorem, we assume now that  $x^{(1)}$  is degenerate local minimum and  $x^{(2)}$  is non-degenerate local minimum. In this case we only have  $D_{y^{(2)}}^2 E(x_\epsilon^{(2)}) > \delta$ . Nevertheless, in Appendix B we show that  $D_{y^{(1)}}^2 E(x_\epsilon^{(1)}) - D_{y^{(1)}}^2 E(x^{(1)})$  is of order  $\epsilon \log \epsilon^2$ . Therefore, since  $D_{y^{(1)}}^2 E(x^{(1)}) \geq 0$ , it follows that we can choose  $\epsilon$  small enough such that  $D_{y^{(1)}}^2 E(x_\epsilon^{(1)}) > -\delta/2$ .

As pointed out in the proof of the non-singular case of theorem 5, the only  $y \in x^\perp$  (recall  $x \equiv x^{(1)} \otimes x^{(2)}$ ) for which it is possible to have  $D_y^2 E(x) = 0$  is  $y = y^{(1)} \otimes x^{(2)}$ . However, the equality in Eq. (41) implies that  $x$  is a local minimum in this direction and this is also true even if  $x^{(i)}$  are singular. We will therefore assume now that  $y$  is not of the form  $y^{(1)} \otimes x^{(2)}$ . By following precisely the same steps of the proof of Theorem 5 (in the non-singular case) we get that for all other  $y \in x^\perp$  we have  $D_y^2 E(x_\epsilon) > \delta - \delta/2 = \delta/2$ . We therefore get  $D_y^2 E(x) > 0$  in the limit  $\epsilon \rightarrow 0$ . This completes the proof of the second part of theorem 5.

## V. DISCUSSION

We have shown that the minimum entropy output of a quantum channel is locally additive (assuming at least one of the two local minima is non-degenerate). Our proof consists of two key ingredients. The first one is the use of the divided difference approach, which enabled us to calculate directional derivatives explicitly, and the second one is the explicit use of the complex structure. In the appendix B of [10] we show that there exists counterexamples for local additivity over the real numbers. These counterexamples precludes the existence of a more straightforward differentiation argument than the complex structure based argument given here.

The fact that the minimum entropy output is not globally additive makes local additivity of even greater interest to quantum information theorists. It suggests that it is some global feature, of the quantum channels involved, that corresponds to cases of non-additivity of the minimum entropy output. Perhaps one way to improve our understanding in this direction is to study properties of generic channels. In particular, it seems quite possible to us that for generic channels (or generic subspaces) the entropy output have a *finite* number of isolated non-degenerate critical points.

*Acknowledgments:*— We acknowledge many fruitful discussions with A. Roy and J. Yard in the earlier stages of this work. GG research is supported by NSERC. The authors acknowledge support from PIMS CRG MQI, MITACS, and iCore for Shmuel Friedland's visits to IQIS in Calgary.

### Appendix A: Proof of Theorem 7

*Proof.* Let  $\lambda_1(t) \geq \dots \geq \lambda_n(t) \geq 0$ , for  $t > 0$ , be the eigenvalues of  $\rho(t)$ . Rellich's theorem yields that each  $\lambda_i(t)$  is analytic in  $t$  in a neighbourhood of  $t = 0$ . So  $\lambda_i(0) = \lambda_i(\rho) > 0$  for  $i = 1, \dots, r$  and  $\lambda_i(0) = 0$  for  $i = r + 1, \dots, n$ . Since each  $\lambda_i(t) \geq 0$  it follows that the Taylor expansion of each  $\lambda_i(t) \not\equiv 0$ , for  $i > r$ , must start with  $t$  to a positive even power times a positive constant. I.e.  $\lambda_i(t) = \lambda_{i,2n_i} t^{2n_i} (1 + O(t))$ , where  $\lambda_{i,2n_i} > 0$  and  $n_i$  is a positive integer for  $i > r$ . This shows that  $S(\rho(t)) = -\sum_{i=1}^n \lambda_i(t) \log \lambda_i(t)$  must be of the form (43). Furthermore,  $K = 0$  if and only if  $n_i \geq 2$  for all  $i > r$ . So if  $K = 0$  and not all  $\lambda_i(t)$  are identically zero for  $i > r$ , then  $k = \min\{n_i - 1, \lambda_{i,2n_i} > 0\}$ .

It is left to show that  $K = \text{Tr}(y_{22}y_{22}^*)$ . Let  $X = \begin{bmatrix} 0 & x \\ x^* & 0 \end{bmatrix}, Y = \begin{bmatrix} 0 & y \\ y^* & 0 \end{bmatrix}$ . Recall that the pencil  $X + tY$  has  $n$  nonnegative and  $n$  nonpositive eigenvalues

$$\sigma_1(t) \geq \dots \geq \sigma_n(t) \geq 0 \geq -\sigma_n(t) \geq \dots \geq -\sigma_1(t).$$

The singular values of  $x + ty$  are the  $n$  nonnegative eigenvalues of  $X + tY$ . Hence, the eigenvalues of  $\rho(t)$  are  $\frac{\sigma_i(t)^2}{1+t^2}$  for  $i = 1, \dots, n$ . Let  $\sigma_i(t) = \sigma_{i,1}t + O(t^2)$  for  $t > 0$  and  $i > r$ . Hence the coefficient of  $t^2$  in the  $i$ -th eigenvalue of  $\rho(t)$ , for  $i > r$ , is  $\sigma_{i,1}^2$ . Thus  $K = \sum_{i=r+1}^n \sigma_{i,1}^2$ .

Let  $P \in \mathbb{C}^{2n \times 2n}$  be the orthogonal projection on the zero eigenspace of  $X$ . Then  $PYP((I - P)\mathbb{C}^{2n}) = \mathbf{0}$ . The other possible nonzero eigenvalues of  $PYP$  are  $\sigma_{r+1,1} \geq \dots \geq \sigma_{n,1} \geq 0 \geq -\sigma_{n,1} \geq \dots \geq -\sigma_{r+1,1}$ , which are the eigenvalues of the restriction of  $PYP$  on the kernel of  $\rho$  [8, 18] or [9, §3.8]. The restriction of  $PYP$  to the kernel of  $X$  is  $\begin{bmatrix} 0 & y_{22} \\ y_{22}^* & 0 \end{bmatrix}$ , obtained by deleting the corresponding rows and the columns in  $Y$ . Hence

$$2K = 2 \sum_{i=r+1}^n \sigma_{i,1}^2 = \text{Tr}((PYP)^2) = \text{Tr}(y_{22}y_{22}^* + y_{22}^*y_{22}).$$

This completes the proof. □

### Appendix B: Formula for the second derivative in the singular case

**Proposition 10.** *Let*

$$x_\epsilon = \begin{bmatrix} x_{11} & 0_{r,n-r} \\ 0_{n-r,r} & \epsilon I_{n-r,n-r} \end{bmatrix} \quad \text{and} \quad y = \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & 0_{n-r,n-r} \end{bmatrix}, \quad (\text{B1})$$

where  $0_{i,j}$  are  $i \times j$  zero matrices, and  $x_{11}, y_{11} \in \mathbb{C}^{r \times r}$ ,  $y_{12} \in \mathbb{C}^{r \times n}$ ,  $y_{21} \in \mathbb{C}^{n \times r}$ . We also assume that  $x_{11} = \text{diag}\{\sqrt{p_1}, \dots, \sqrt{p_r}\}$  is non singular. Then, the limit of  $D_y^2 E(x_\epsilon)$  when  $\epsilon$  goes to zero exists and equals to

$$\lim_{\epsilon \rightarrow 0} D_y^2 E(x_\epsilon) = D_{y_{11}}^2 E(x_{11}) - 2\text{Tr}[(y_{12}y_{12}^* + y_{21}^*y_{21}) \log \rho_{11}], \quad (\text{B2})$$

where  $\rho_{11} \equiv x_{11}x_{11}^*$ .

*Remark.* The contribution of the normalization factor of  $x_\epsilon$  is of order  $O(\epsilon^2)$  and therefore ignored here.

*Proof.* The proof is based on a straightforward calculation. The expression for the second derivative given in Eq.(17), can be written as:

$$D_y^2 E(x_\epsilon) \equiv \frac{d^2}{dt^2} S(\rho_\epsilon(t)) \Big|_{t=0} = -2 \left( S(\rho_\epsilon) + \sum_{j=1}^n \sum_{k=1}^n G_{jk} \right),$$

where  $\rho_\epsilon \equiv x_\epsilon x_\epsilon^*$ ,  $\gamma_0 \equiv x_\epsilon y^* + y x_\epsilon^*$

$$G_{jk} \equiv |y_{jk}|^2 \log p_j + \frac{\log p_j - \log p_k}{2(p_j - p_k)} |(\gamma_0)_{jk}|^2. \quad (\text{B3})$$

If both  $j, k$  are smaller or equal to  $r$ , then clearly those  $G_{jk}$  terms contribute to  $D_{y_{11}}^2 E(x_{11})$ . Now, if both  $j > r$  and  $k > r$  then  $y_{jk} = 0$  and we have  $G_{jk} = 0$ . Hence, we get

$$D_y^2 E(x_\epsilon) = D_{y_{11}}^2 E(x_{11}) - 2 \sum_{j=r+1}^n \sum_{k=1}^r (G_{jk} + G_{kj}) \quad (\text{B4})$$

We therefore focus now on the expressions for  $G_{jk}$  and  $G_{kj}$  in the case  $j > r$  and  $k \leq r$ .

Writing  $x_\epsilon = \text{diag}\{\sqrt{p_1}, \dots, \sqrt{p_n}\}$  with  $p_j = \epsilon^2$  for  $j > r$  we have

$$\begin{aligned} (\gamma_0)_{jk} &= \sqrt{p_j} \bar{y}_{kj} + \sqrt{p_k} y_{jk} = \sqrt{p_k} y_{jk} + O(\epsilon) \\ (\gamma_0)_{kj} &= \sqrt{p_k} \bar{y}_{jk} + \sqrt{p_j} y_{kj} = \sqrt{p_k} \bar{y}_{jk} + O(\epsilon), \end{aligned}$$

where the last equality was obtained by setting  $p_j = \epsilon^2$ . We therefore have  $|(\gamma_0)_{jk}|^2 = |(\gamma_0)_{kj}|^2$  up to  $O(\epsilon)$ . From the expressions above we get for  $j > r$  and  $k \leq r$  the following formulas:

$$\begin{aligned} G_{jk} &= |y_{jk}|^2 \log p_j + \frac{\log p_j - \log p_k}{2(p_j - p_k)} (p_k |y_{jk}|^2 + O(\epsilon)) \\ G_{kj} &= |y_{kj}|^2 \log p_k + \frac{\log p_k - \log p_j}{2(p_k - p_j)} (p_k |y_{jk}|^2 + O(\epsilon)) \end{aligned}$$

Since  $p_j = \epsilon^2$  and  $p_k > 0$ , we have

$$\begin{aligned} G_{jk} &= |y_{jk}|^2 \log \epsilon^2 + \frac{\log \epsilon^2 - \log p_k}{2(\epsilon^2 - p_k)} (p_k |y_{jk}|^2 + O(\epsilon)) = \frac{1}{2} |y_{jk}|^2 \log \epsilon^2 + \frac{1}{2} |y_{jk}|^2 \log p_k + O(\epsilon \log \epsilon) \\ G_{kj} &= |y_{kj}|^2 \log p_k + \frac{\log p_k - \log \epsilon^2}{2(p_k - \epsilon^2)} (p_k |y_{jk}|^2 + O(\epsilon)) = |y_{kj}|^2 \log p_k + \frac{1}{2} |y_{jk}|^2 \log p_k - \frac{1}{2} |y_{jk}|^2 \log \epsilon^2 + O(\epsilon \log \epsilon) \end{aligned}$$

Hence,

$$G_{jk} + G_{kj} = |y_{jk}|^2 \log p_k + |y_{kj}|^2 \log p_k + O(\epsilon \log \epsilon).$$

By substituting this expression into Eq.(B4) we get (B2). This completes the proof.  $\square$

- [1] K. M. R. Audenaert and S. L. Braunstein, Comm. Math. Phys., **246**, 443 (2004).
- [2] G. G. Amosov, A. S. Holevo, and R. F. Werner, Problems in Inf. Trans. **36**, 305 (2000).
- [3] C. H. Bennett, D. P. DiVincenzo, T. Mor, P. W. Shor, J. A. Smolin and B. M. Terhal, Phys. Rev. Lett. **82**, 5385 (1999).
- [4] F. G.S.L. Brandao, J. Eisert, M. Horodecki, and D. Yang, Phys. Rev. Lett. **106**, in press (2011). [quant-ph/1010.5074].
- [5] F. G.S.L. Brandao and M. Horodecki, Open Syst. Inf. Dyn. **17**, 31 (2010).
- [6] H. Dersken, S. Friedland, G. Gour, D. Gross, L. Gurvits, A. Roy, and J. Yard. On minimum entropy output and the additivity conjecture. Notes of Quantum Information Group, American Institute for Mathematics workshop "Geometry and representation theory of tensors for computer science, statistics and other areas", July 21-25, 2008.



- [7] M. Fannes, B. Haegeman, M. Mosonyi, and D. Vanpeteghem, quant-ph/0410195.
- [8] Extremal eigenvalue problems, *Bull. Brazilian Math. Soc.* 9 (1978), 13-40.
- [9] S. Friedland, *Matrices*, A draft of a book in preparation, <http://homepages.math.uic.edu/~friedlan/bookm.pdf>
- [10] S. Friedland, G. Gour, A. Roy, Local extrema of entropy functions under tensor products, *Quantum Information and Computation*, Vol.11 No.11-12 pp1028-1044, eprint: math-ph/1105.5380.
- [11] M. Fukuda, C. King, and D. K. Moser, *Commun. Math. Phys.* **296**, 111-143, 2010.
- [12] G. Gour and S. Friedland, arXiv:1105.6122 [quant-ph].
- [13] G. Gour and N. Wallach, *Phys. Rev. A* **76**, 042309 (2007).
- [14] M. B. Hastings, *Nature Physics* **5**, 255 (2009).
- [15] A. S. Holevo, The additivity problem in quantum information theory. In *International Congress of Mathematicians. Vol. III*, pages 999–1018. Eur. Math. Soc., Zürich, 2006.
- [16] R.A. Horn and C.R. Johnson, *Topics in Matrix Analysis*, Cambridge Univ. Press, 1999.
- [17] R. Horodecki, P. Horodecki, and M. Horodecki, *Rev. Mod. Phys.* **81**, 865 (2009).
- [18] T. Kato, *Perturbation Theory for Linear Operators*, Springer-Verlag, 2nd ed., New York 1980.
- [19] C. King, *J. Math. Phys.* **43**, 4641 (2002).
- [20] C. King, *Quant. Inf. Comp.* **3**, 186 (2003).
- [21] C. King and M. B. Ruskai, *IEEE Trans. Inf. Theory* **47**, 192 (2001).
- [22] M. Ohya and D. Petz, *Quantum entropy and its use*, Springer (2004).
- [23] D. Petz, *Quantum information theory and quantum statistics*, Springer (2008).
- [24] M. E. Shirokov, *Problems of Information Transmission*, **42**, 23-40 (2006).
- [25] P. W. Shor, *J. Math. Phys.* **43**, 4334 (2002).
- [26] P. W. Shor. Equivalence of additivity questions in quantum information theory. *Comm. Math. Phys.*, 246(3):453–472, 2004.
- [27] G. Smith and J. Yard, *Science* **321**, 1812 (2008).
- [28] J. Yard, private communication.
- [29] In [6, 10] it was shown to be true for a large class of functions (not only for the von-Neumann entropy) including all the  $p$ -norm entropy functions.